Chapter 4

DISCUSSION

1. Two *hmgs* genes encoding HMG-CoA synthase in *H. brasiliensis*

The titration of rubber latex UniZAP-XR cDNA library phage prepared as described by Suwanmanee et al. 2002, and performed by infecting in the bacteria XL1-Blue MRF' strain, exhibited the plaque clear zone on the plate and was calculated to be $3.5x10^5$ pfu/ml. This was not a high yield infection; however, it could be used for screening the new *hmgs*. The primary screening using *H. brasiliensis hmgs1* as a probe showed five positive clones and, after secondary screening to confirm the insertion of *hmgs* cDNA, all five clones showed positive signals in the tertiary screening and the DNA sequences of the insert cDNA was determined. Only one clone contained a new partial cDNA, *hmgs2*. After 5' RACE was performed, the 1,916 bp full-length cDNA encoding HMG-CoA synthase 2 in *H. brasiliensis* was obtained.

The nucleotide sequence in the open reading frame of *hmgs2* differs from that of *hmgs1* at 77 positions (Table 12). Of all the changes, 44 nucleotides were silent changes and 33 were missense. As a consequence, HMG-CoA synthase 2 differs from HMG-CoA synthase 1 in 28 amino acids (Figure 20 and Table 8). In all of the 77 bases changes, 37 bp were observed only one time, 39 bp were observed twice, and 1 bp was observed three times. The changes at these nucleotides were observed on the independent occasions which rules out sequencing errors and random error in the cloning and strongly indicates that *hmgs2* is a new gene for HMG-CoA synthase in *H. brasiliensis*. Like the *hmgs1*, the isolated *hmgs2* cDNA encodes a cytosolic HMG-CoA synthase because it lacks the N-terminal mitochondrial leader peptide sequence

that makes up the first 37 amino acids of the human mitochondrial HMG-CoA synthase which marks the enzyme for import into mitochondria.

Table 12. Nucleotide sequence differences between *H. brasiliensis* *hmgs1* and *hmgs2*

There are 1,392 comparable positions
77 nucleotide differences (8%)
15 involved codon position 1
10 involved codon position 2
52 involved codon position 3

The fact that *H. brasiliensis* HMG-CoA synthase is encoded by two *hmgs* genes, is also supported by the presence of two HMG-CoA synthase in mammals, the mitochondrial and cytosolic isoforms as reported by Ayte et al. (1990) and HMG-CoA synthase 1 and 2 in cytoplasm of *B.germanica* (Buesa et al. 1994). In other plants, up to four *hmgs* cDNA are found in *Brassica juncea*; *Bjhmgs1*, *Bjhmgs2*, *Bjhmgs3*, and *Bjhmgs4*. *Bjhmgs1* shows HMG-CoA synthase activity that is inhibited by F244, a specific inhibitor of HMG-CoA synthase. *Bjhmgs1* has amino acid sequence identity of 97%, 96%, and 95% with *Bjhmgs2*, *3*, and *4*, respectively. *Bjhmgs1* also has a nucleotide identity with *Bjhmgs2, 3*, and *4* of 85-92% (Alex et al. 2000). In addition,

southern blot analysis revealed gene families encoding HMG-CoA synthase in *B. juncea* and a summation of homologous genes in the fusion amphidiploid genome of *Brassica* species a bi-parental species derived from diploids *B. nigra* and *B. campestris*. Moreover, Montamat et al. (1995) suggested the possibility of a second gene encoding HMG-CoA synthase on the basis of Southern blot analysis of *A.thaliana* DNA. Wegener et al. (1997) detected two putative *Pinus sylvestris* HMG-CoA synthase mRNAs. Therefore, it appears that there are gene families encoding HMG-CoA synthase in plants, as is often true for HMG-CoA reductase. So far from the cDNA library screening and 5' RACE experiment we found only *hmgs1* and *hmgs2* gene from the DNA sequencing of positive clones.This is in line with the suggestion that *H. brasiliensis* has at least two genes encoding for HMG-CoA synthase by Suwanmanee et al. (2002). The identification of the second gene encoding HMG-CoA synthase in *H. brasiliensis* was also supported by the studies on the amino acid sequence alignment and phylogenetic tree analysis.

2. Tissue expression of *H. brasiliensis hmgs2* cDNA

The mRNA levels of *hmgs2* in *H. brasiliensis*, detected by semiquantitative RT-PCR, were normalized by comparison with the 18S rRNA standard transcript and the amplification was up to 40 cycles for the *hmgs2* specific primers. The use of 18S rRNA is recommended as an internal standard for mRNA quantification study; because mRNA variations are weak in comparison and cannot highly modify the total RNA level (Thellin et al. 1999). The *hmgs2* mRNA levels from various tissues such as latex, petiole, and leaf, were different (Figure 22). The result is consistent with tissue specific expression of the *hmgs1* in the previously study of Suwanmanee et al. (2002).

Similarly, mitochondrial *hmgs* showed tissue specific expression in rat liver and intestine (Thumelin et al. 1993). The cytosolic *hmgs1* and *hmgs2* in *B. germanica* are differently expressed throughout the developmental stages (Martinez-Gonzales et al. 1993).

In the present study, an attempt has been made to compare the mRNA expression of *hmgs1* and *hmgs2* by RT-PCR. The amplification of *hmgs1* fragment was performed by using the total RNA from latex, petiole, and leaf as used for the amplification of *hmgs2* and two primers specific for *hmgs1*. Agarose gel electrophoresis showed higher intensity of *hmgs1* band (data not shown). The result suggested the possibility that *hmgs1* and *hmgs2* are equally expressed. This may be due to the high indentity of *hmgs1* and *hmgs2* of 92% and the sequence length of *hmgs1* is shorter than *hmgs2*. Therefore, it is difficult in selecting appropriate primers specific for *hmgs1*.

The expression of both *hmgs1* and *hmgs2* is correlated with the presence of more laticiferous cells in the particular tissue; latex and petiole. This finding is consistent with the report that HMG-CoA reductase genes; *hmg1*, *hmg2*, and *hmg3* are differentially expressed (Chye et al. 1992). The *hmg3* is equally abundant in laticifers and leaves and is possibly involved in the synthesis of other isoprenoids. In contrast, *hmg1* is expressed predominantly in the laticifers, the cells specific to rubber biosynthesis. Since both *hmgs* genes appear to be in laticifers in *H. brasiliensis*, it is likely that both *hmgs1* and *hmgs2* encode the enzyme HMG-CoA synthase that is involved in rubber biosynthesis.

3. Multiple alignment of HMG-CoA synthase and ACP synthase III.

The protein sequence of *Hevea brasiliensis* HMG-CoA synthase and others were aligned by the progressive alignment program (Feng and Doolittle, 1990) which was run under the UNIX operating system environment. Alignments were obtained by a binary alignment algorithm (Needleman & Wunsch, 1970); the similarity scale of each amino acid is based on the mutation matrix (Dayhoff et al. 1978), which generates a score for all amino acid interchanges in a given pair of the sequences. The alignment starts from the closest pair and adds in the rest of sequences progressively. The program was run automatically by checking the next two sequences at each round, priority being determined on the basis of the higher similarity score. This operation was continued until all the sequences were in alignment.

Proper multiple alignments were obtained by careful "cropping" of the sequences to approximately the same lengths to show all the conserved regions that exist in HMG-CoA synthase. A detailed comparison of plant sequences reveals the expected quite high identity to *H. brasiliensis* HMG-CoA synthase (Table 10). This means that there is not much diversity in various known plant HMG-CoA synthases. Lower identity occurs with animals and bacterial HMG-CoA synthase and ACP synthase III.

Among the conserved residues throughout all species shown in the multiple alignment of all HMG-CoA synthase and ACP synthases sequences were those corresponding to *H. brasiliensis* HMG-CoA synthase, cysteine-117, histidine-247, glycine-325, asparagines-326, glycine-358, and glycine-360 (Figure 24). The cysteine-117 is known to be an active site residue in both kinds of proteins. Protein chemistry (Vollmer et al. 1988) and mutagenesis studies (Misra et al.1993) of this enzyme from

avian sources allowed the unambiguous functional assignment of cysteine-129 as the active site amino acid that forms the acetyl-*S*-enzyme intermediate. A study on an acetylation of Cys^{112} in β-Ketoacyl-acyl carrier protein synthase III (ACP synthase III) from *E.coli* reported the catalytic role and clearly defined the primer binding pocket of Cys^{112} (Qiu et al. 1999).

Histidine-247 (His^{247} in *Hevea brasiliensis*) is known to play a catalytic role. The evidence for the interaction of avian HMG-CoA synthase histidine-264 with acetoactyl-CoA suggests that the imidazole of His^{264} in the avian enzyme plays a role in binding the secondary substrate, acetoacetyl-CoA, by interacting with the carbonyl oxygen of the thioester functionality (Misra & Miziorko 1996). From the alignment, there was also a totally conserved asparagine residue (Asn^{326} in *H*evea *brasiliensis*). Modeling based on a bound CoA molecule suggests catalytic roles for His^{244} and Asn^{274} in ACP synthase III (Qiu et al. 1999). While a catalytic role of Gly^{325}, Gly^{358}, and Gly^{360} has not yet been specified in the two enzymes, among these residues, Cys^{117}, His^{247}, and Asn^{326} are absolutely invariant in all available deduced HMG-CoA synthase sequences and ACP synthase III sequences, indicating that $Cys^{117}$and His^{247} are required for formation of enzyme-substrate complex and Asn^{326} is involved in catalytic reaction for HMG-CoA synthase.

Inspection of the structure of ACP synthase III (1HZP; PDB file) showed that the conserved Asn (Asn^{274} in *M. tuberculosis*) was very near the active site Cys^{112} and His^{244}. Therefore Asn^{326} in *H. brasiliensis* is in position to play a catalytic role in HMG-CoA synthase as in the ACP synthase III protein. In fact, the catalytic role should be the same, even though these two enzymes are of two different classes.

HMG-CoA synthase is an oxo-acid-lyase, whereas ACP synthase III is an acyltransferase. However, they both are condensing enzymes.

Three condensing enzymes are involved in bacterial fatty acid biosynthesis: β-ketoacyl-ACP synthase I, II, and III, respectively. Each catalyzes a distinct biochemical reaction. The ACP synthase III (EC 2.3.1.41) is the key condensing enzyme in the initiating condensation reaction in bacterial fatty acid biosynthesis pathway, catalyzing the condensation of acetyl-CoA with malonyl-ACP to yield acetoacetyl-ACP (Magnuson et al.1993). By writing the exact mechanism of actions of the two enzymes, HMG-CoA synthase and ACP synthase III activity in detail, one can see its mechanism, and how one enzyme adapted to a new machanism. The reaction of ACP synthase III occurs in three steps as in Figure 34 (Abbadi et al. 2000). First, the acetyl is transferred from acetyl-CoA to a catalytic cysteine of enzyme to become an acetylated-enzyme, then the decarboxylation of malonyl-ACP forms a carbanion, followed by condensation of the acetyl group and carbanion to give a β-ketoester (β-ketobutyryl-ACP, acetoacetyl-ACP) (Magnuson et al.1993).

The mechanism of HMG-CoA synthase in the reactions also involves three steps as shown in Figure 35 (Chun et al. 2000a): acetylation of enzyme at cysteine, condensation of the acetyl group with acetoacetyl-CoA to form acetoacetyl-CoA-enzyme, and hydrolysis of the acetoacetyl-CoA-enzyme to yield HMG-CoA and releasing the enzyme. Both HMG-CoA synthase and ACP synthase III catalyse the condensation reaction. The mechanism of ACP synthase III was adapted by using malonyl-ACP as a substrate after the enzyme was acetylated by acetyl-CoA, while acetyl-S-enzyme was condensed with acetoacetyl-CoA to form HMG-CoA-S-enzyme (condensation intermediate) and yield the product, HMG-CoA for HMG-CoA

synthase. On the other hand, acetyl-S-enzyme was condensed with malonyl-ACP to yield β-ketobutyryl-ACP in ACP synthase III. However, they catalyze the reaction in the same way, following a Ping Pong mechanism in which after binding the first substrate (acetyl-CoA) a first product (reduced CoA) is released before binding the second substrate as in an acid-base catalysis (Abbadi et al. 2000).

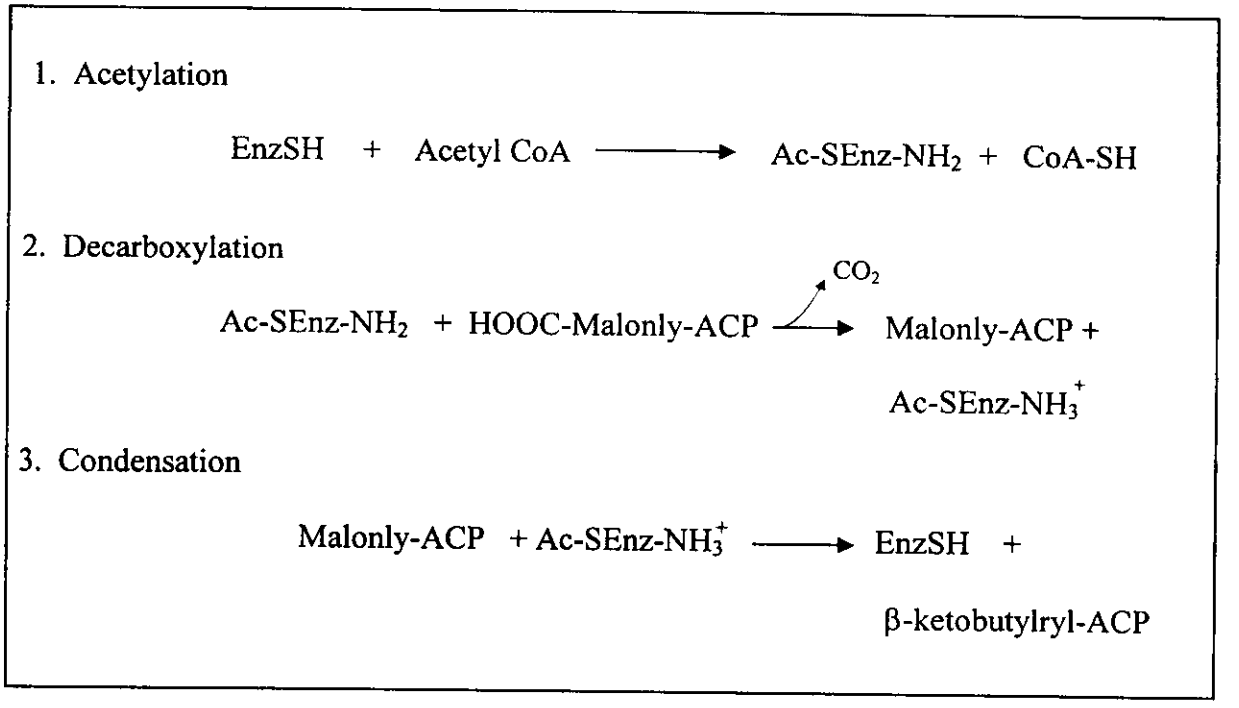

Figure 34. Reactions catalyzed by ACP synthase III (Abbadi et al. 2000).

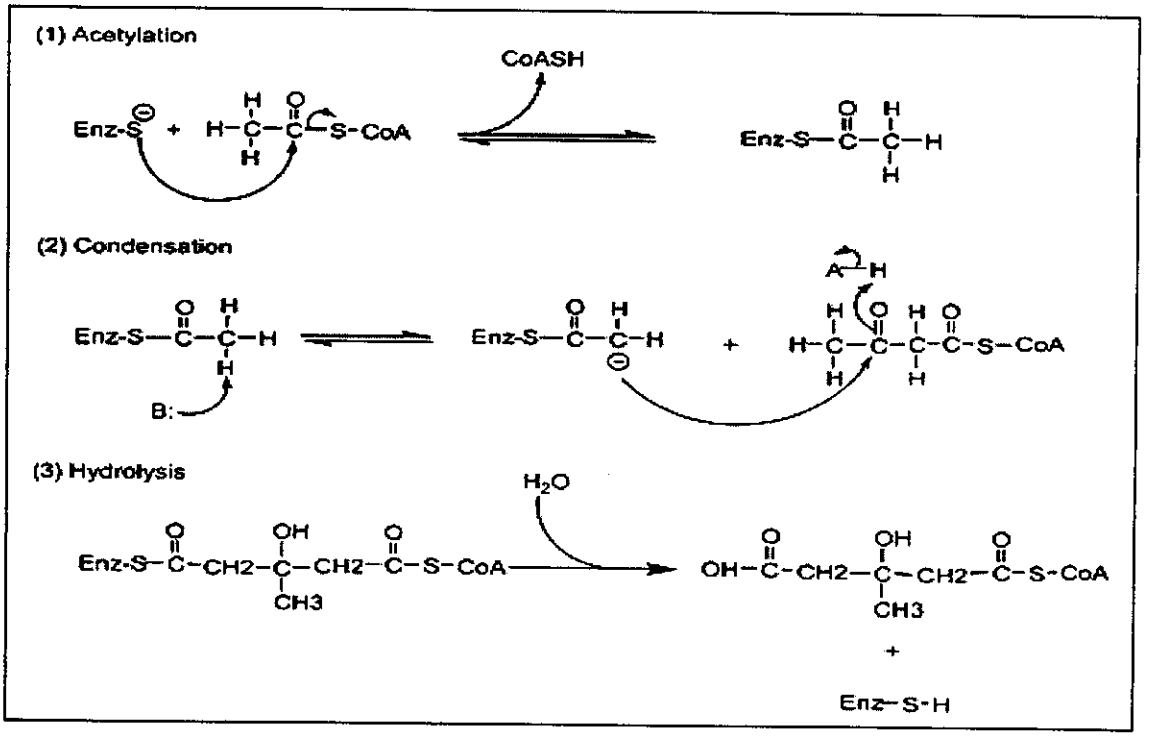

Figure 35. Reactions catalyzed by HMG-CoA synthase (Chun et al. 2000a).

4. Phylogenetic analysis

It is an axiom of modern biology that all extant organisms are descended from a common ancester (Darwin, 1859). A vital component of this evolutionary history is the passage of genetic information from parent to offspring. All of life is genetically connected via a vast phylogenetic tree

A phylogenetic tree is a diagram showing the evolutionary interrelation of a group of organisms that originated from a shared ancestor. The tree gives the "branching order" or "topology", which is a reflection of the gene duplication giving rise to new gene products of divergence for organisms that share a common ancester. The branch lengths are indicative of the evolutionary distance separating the sequences that existed when these events occured in the present time (Doolittle, 1987). The ancestor is at the tree root; organisms that have arisen from it are placed at the ends of tree branches. The distance of one group from another group indicates the degree of relationship between them; closely related groups are located on branches close to each other. Phylogenetic trees can be either gene trees or species trees or a combination. The gene tree represents the phylogeny of a set of current genes, while the species tree represents the phylogeny of the current genes organized by their host-species.

One of the most important phenomena that leads to the evolution of new species is the duplication of genes, which can create extra copies of a gene. A large proportion of genes are apparently the product of gene duplication. Evolution proceeds via gene duplication and modification; it is through their successive application that Nature has created the vast diversity of current genes. A gene encoding an essential function may be duplicated, one copy preserving the function for its organism.

The other, however, is free to be modified, using its ancestral gene as its starting point. Knowing when ancestral gene duplications occurred is indispensable for reconstructing reliable phylogenetic relationships, which in turn allows one to refine the prediction of a genetic function. For example, globin genes have duplicated and diverged. Leghemoglobin binds oxygen, playing an essential role in nitrogen fixation by legumes and other plants, while phycocynanin, another protein, is a photosynthetic pigment (Donoghue, 1992).

In this study, a phylogenetic tree constructed from the result of proper multiple alignments (Feng & Doolittle, 1990) is a combination of a gene tree and a species tree. The gene tree represented the group of the topology of HMG-CoA synthase genes and ACP synthase III genes in each species. The tree was divided into two major groups, HMG-CoA synthase and ACP synthase III, reflective of the multiple alignment of 30 amino acid sequences. The branching order is species tree that represents the evolutionary pathway of HMG-CoA synthase genes in several species among vertebrates, insects, plants, yeast, worm, and bacteria, and of ACP synthase III gene in archeabacteria and eubacteria. Phylogenetic tree shows HMG-CoA synthase 1 and HMG-CoA synthase 2 were recently diverged by gene duplication in terms of amino acid replacement which was subjected to positive selection. Gene duplication is important for generating new copies of genes of the same function, thereby enabling the production of large quantity of RNA or proteins (Li, 1997). Likewise, the two condensing enzymes, HMG-CoA synthase and ACP synthase III have emerged by gene duplication from their common ancestor gene in the past.

The evolutionary rates of ACP synthase III genes appear to be slower than HMG-CoA synthase after divergence. In the case of HMG-CoA synthase, the average

branch lengths from the divergence point to the extant species appear not to differ significantly between different species, suggesting similar evolutionary rates for HMG-CoA synthase on average; also in the ACP synthase III, each species did not change faster than the others. However, the evolutionary rate of ACP synthase III gene after diverging from HMG-CoA synthase by gene duplication event is slower than that of HMG-CoA synthase.

All known HMG-CoA synthase amino acid sequences in mitochondria and cytoplasm showed a high degree of conservation near the N-terminal which declines toward the C-terminal (Boukaftane et al. 1994). In the phylogenetic tree, the tree branches among species of mitochondrial and cytosolic enzyme are close to each other. This result suggests that the two isoenzymes diverged from a common ancestor gene in the past with gene duplication and then they became paralogous genes. This event occurred after the speciation event to the ancestors of insects and mammals and they evolved at the same rate in the two groups.

In the evolutionary lineage, *hmgs* was duplicated to a new gene in bacteria first; this is also true in the case of ACP synthase III genes. Consequently, the bacterial amino acid sequences of ACP synthase III obtained by using HMG-CoA synthase as a query in the GenBank database enable us to align and determine the degree of identity to HMG-CoA synthase. The evolutionary lineage of *hmgs* then continues among eukaryotes, *C. elegans* and yeast, respectively; after that, this gene diverged to plants, insects, bird, and mammals. This finding supports the postulate that HMG-CoA synthase belongs to a large protein family comprising other acetyl-CoA condensing enzymes, such as ACP synthase III of fatty acid biosynthesis and chalcone synthase of plant phenylpropanoid metabolism (Lange et al. 2000). The topology of the tree shows

that all the ACP synthase III were not grouped together in their own cluster. The phylogenetic position of the archaebacterium ACP synthase III is related to HMG-CoA synthase proteins more than to eubacterium ACP synthase III. Gene duplications are traditionally considered to be a major evolutionary source of new protein and produce two functionally redundant, paralogous genes, which thereby frees one of them from selective constrains. This unconstrained paralog is then free to accumulate neutral mutations that would have been deleterious in a unique gene.

The phylogenetic tree of HMG-CoA synthase and ACP synthase III shows that both proteins were derived from duplication of a gene involved in condensation. Gene duplications leading to a new function are proceded by a period of gene sharing such that the original, duplicated gene, encodes two distinct functions, likely diverging to two paralogous genes, the HMG-CoA synthase gene and ACP synthase III in this study. With regard to the structure of two proteins that were separated by gene duplication, residues involved in catalysis and internal structural interactions are likely to be conserved more than residues that are not. Therefore, in this study the X-ray crystallograhpy structure of ACP synthase III was used to predict the model of HMG-CoA synthase.

5. Prediction of the possible secondary structure of HMG-CoA synthase

The prediction of protein secondary structure is the most general method of obtaining some structural information from any newly-determined sequence. Secondary structure prediction is important in establishing alignments during model building by homology and also the first step in attempting to generate tertiary structure models by docking α-helices and β-sheets. There exist a bewildering variety of

methods for predicting the secondary structure of proteins from primary structure. The statistical methods are based on studies of the database of proteins of known primary and secondary structure, such as the Chou-Fasman and GOR methods. The Chou-Fasman method (Chou and Fasman, 1974) is a statistical method based on the calculation of the statistical propensities of each residue to be found in either an α-helix or β-sheets. Development of this program took 12 known X-ray structures of unrelated proteins and examined them carefully to see which amino acids were in the α- helices, β-sheets, and the turns. Then a program was designed which looked at neighboring amino acids. The designations were then used to find areas of probable α-helices or β-sheets in the protein sequence to be predicted. Probable areas of α-helix and β-sheets were then modified by a series of rules to produce the final prediction. The GOR (Garnier-Osguthorpe-Robson) method is based on the idea of treating the primary sequence and the sequence of secondary structure as two messages related by translation process, this translation process is then examined using information theory.

In this study, the 3-D dimension structure derived from X-ray crystallography of ACP synthase III which is related to HMG-CoA synthase was used as template to make the structural alignment, and then the secondary structure of HMG-CoA synthase was predicted. From the Protein Data Base (PDB), 3-D dimension structures have been reported for two ACP synthases III from *E. coli* and *M. tuberculosis*, so one of them, ACP synthase III from *M. tuberculosis*; 1HZP (Scarsdale et al. 2001), was used.

There is now ample evidence to realize that gene duplication is the most important mechanism for generating new genes and new biochemical processes selected by nature, which have facilitated the evolution of complex organisms (Li,

1997). Furthermore, internal, (partial) gene duplication plays a major role in increasing the functional complexity of genes in the course of evolution. Many proteins of present-day organisms show internal repeats of amino acid sequences, and the repeats often correspond to the functional or structural domains of the proteins (Barker et al. 1987). The interior parts of diverged proteins generally change more slowely than the exterior parts of those proteins. As it happens, both HMG-CoA synthase and ACP synthase III enzymes still retain the three important active residues, Cys^{117}, His^{247}, and Asn^{326} inside the hydrophobic region and others, the function of which has not yet been revealed. Prediction of the possible secondary structure of HMG-CoA synthase from PDB information of its relative enzyme led us to postulate the position of important residues, and it is almost certain the secondary structure of HMG-CoA synthase will look like ACP synthase III.

6. Expression of wild type and mutant *hmgs1* in E. coli

In order to study the catalytic role of Asn^{326} in *H. brasiliensis* HMG-CoA synthase, the wild type *hmgs1* and mutant gene at the nucleotide positions translated to Cys^{117} and Asn^{326} were substituted by nucleotides translated to Ala by point mutations. The mutant *hmgs1* genes were characterized by DNA sequencing with 5' flanking primer for C117A and 3' flanking primer for N326A. The expression conditions were optimized to yield high expression (Figure 27 and 28). The recombinant proteins were produced in soluble form and showed molecular masses close to the purified HMG-CoA synthase from C-serum of rubber latex (Figure 29). The *H. brasiliensis hmgs1* expressed in *E. coli* exhibited the enzyme activity; however, the activity of the recombinant enzyme in crude extract from the wild type was rapidly lost, similar to

the recombinant *Brassica juncea* HMG-CoA synthase (Alex et al. 2000). This contrasts with crude HMG-CoA synthase in C-serum, which is stable at -70°C for a longer period (Suvachittanont & Wititsuwannakul, 1995). This is most likely because the presence of negatively interfering activities acting on the acetyl-CoA and acetoacetyl-CoA, probably thioesterase that attack the CoA moiety and thus render the substrate inactive for HMG-CoA synthase (Van der Heijden et al. 1994a and b). Although only low quantities of recombinant *H. brasiliensis* HMG-CoA synthase were produced after IPTG induction, the enzyme was sufficient for HMG-CoA synthase assay. Induction of *hmgs* expression by IPTG increases the protein accumulation proportionally to the IPTG concentration from 50 μM to 500 μM, higher IPTG concentration at 1 mM final concentration did not increase the induced protein.

It was found that an enzyme assay from the crude extract in the range of our assay condition showed that the activity of mutated HMG-CoA synthase is lower than the wild type enzyme in all cases (Figure 30). The specific activity of the recombinant proteins, support the catalytic role of Cys^{117} and Asn^{326}. The result supports the fact that, an invariant Cys^{117} is required for the catalytic role of HMG-CoA synthase in the same way as those previously reported (Vollmer et al. 1988 and Misra et al. 1993). From the present study, the role of a conserved Asn^{326} in catalytic activity of HMG-CoA synthase was first established. It was found that the Asn^{326} mutant had a distinctly negative effect on the catalytic activity, this result supports that Asn^{326} is an important amino acid required for HMG-CoA synthase, structure folding or activity, as has been reported for its relative, ACP synthase III, another condensing enzyme.

The conserved residues, Cys^{117}, His^{247}, and Asn^{326} (the numbers refer to residues in *H. brasiliensis* HMG-CoA synthase) have been recently demonstrated as

catalytic residues for condensing enzymes (Price et al.2001). In particular, ACP synthase III, the most divergent member of the condensing enzymes family which has been extensively studied for drug targets at catalytic residues in *M. tuberculosis* ACP synthase III, Cys^{122}, His^{244}, and Asn^{274} are conserved (Scarsdal et al. 2001). Moreover, acetylation of Cys^{112} proves that it is clearly defined in the substrate binding pocket. Modeling based on a bound CoA molecule suggests a catalytic role for His^{244}, and Asn^{274} in *E. coli* ACP synthase III (Qiu et al. 1999).

Early studies on HMG-CoA synthase presenting in yeast and avian mitochondrial HMGS have established that the catalytic mechanism of the enzyme involves a three-step process: acetylation, condensation, and hydrolysis (Miziorko et al. 1975). That Cys^{129} is required to form the acetyl-S-enzyme intermediate (Misra et al. 1993 and Vollmer et al. 1988) was confirmed by mutagenesis studies on the recombinant avian enzyme. The His^{264} has been implicated in binding the second substrate, acetoacetyl-CoA, by the kinetic characterization of mutants (Misra & Miziorko 1996). Glu^{95} was proposed to have an additional role in the C-C bond formation upon condensation with the second substrate and also to play a role in the acid/base catalysis (Chun et al. 2000b). Recent work in chicken HMG-CoA synthase, studying influence of several conserved aromatic residues near the active site indicated that Tyr^{130}, Phe^{204}, or Tyr^{376} have an effect on the conformation of the covalent acetyl-S-enzyme intermediate (Misra et al. 2003).